

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-334090

(43) 公開日 平成10年(1998)12月18日

(51) Int.Cl.⁶

識別記号

F I

G 0 6 F 17/22
17/21

G 0 6 F 15/20

5 2 2 L
5 9 6 A

審査請求 有 請求項の数 4 O L (全 9 頁)

(21) 出願番号 特願平9-137215

(22) 出願日 平成9年(1997)5月27日

(71) 出願人 000123262

沖縄日本電気ソフトウェア株式会社
沖縄県那覇市久米2丁目3番15号

(72) 発明者 新里 元

沖縄県那覇市久米2丁目3番15号 沖縄日
本電気ソフトウェア株式会社内

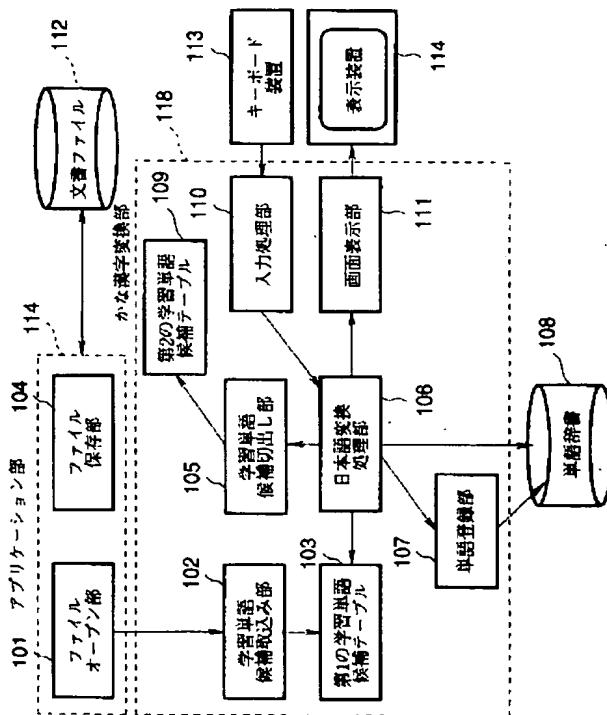
(74) 代理人 弁理士 志賀 正武

(54) 【発明の名称】 日本語入力システム

(57) 【要約】

【課題】 メール文書内の単語辞書に含まれない未登録語を自動的に学習し、その単語を再入力した時に最初の変換候補として出力可能にする。

【解決手段】 ファイル保存部104で未登録およびユーザ登録単語の情報を文書ファイルに付加し、学習単語候補取込み部102はメール受信者がメール文書を開いた時に、文書ファイル内より未登録語及びユーザ登録単語を抽出し、その中で現在使用している単語辞書108に含まれない単語のみを保存して日本語変換処理部106はその保存された読みが入力された場合、第1候補としてこれを出力し、必要なら自動的に単語辞書108へ登録する。



【特許請求の範囲】

【請求項1】 かな漢字変換部に設けられて、利用者が文書内に日本語入力を開始し、確定した文字列の中で、ユーザ登録単語及び単語辞書に存在しない未登録語を抽出する学習単語候補切出し部と、

利用者の入力完了して文書を保存する際、実際に入力された文字列以外に前記学習単語候補切出し部で抽出したユーザ登録単語及び未登録語からなる学習単語情報も併せて文書ファイル内に追加して保存するファイル保存部と、

前記文書ファイルが開かれた場合、そのファイル情報をかな漢字変換部へ通知するファイルオープン部と、

前記かな漢字変換部に設けられて、前記ファイル情報の通知を受けて、文書ファイル内の前記学習単語情報をメモリ上へロードし、その学習単語情報に含まれる単語が現在使用中の単語辞書に存在するか否かを検索し、存在しない未登録の単語のみを取込む学習単語候補取込み部と、

利用者が入力した読み情報が、前記学習単語候補取込み部で取込んだ単語に該当する場合、文法的に問題がなければ最初の変換候補として出力し、更に、その単語が選択された場合は、単語登録するかを利用者に確認し、または必要なら自動的に単語登録を行う日本語変換処理部とを備えたことを特徴とする日本語入力システム。

【請求項2】 前記学習単語候補取込み部が、前記学習単語候補切出し部で切出した単語情報と現在使用中の単語辞書とを照合した結果、未登録語と断定した単語を第1の学習単語候補テーブルに格納することを特徴とする請求項1に記載の日本語入力システム。

【請求項3】 前記かな漢字変換部に、前記学習単語候補取込み部で取込んだ単語を選択して単語辞書へ登録させる単語登録部を有することを特徴とする請求項1に記載の日本語入力システム。

【請求項4】 前記かな漢字変換部に、前記日本語変換処理部で作成した変換結果を表示装置に表示させる画面表示部を有することを特徴とする請求項1に記載の日本語入力システム。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】この発明は、例えば電子メールシステム上でのかな漢字変換処理において未登録語の抽出、学習を行えるようにする日本語入力システムに関する。

【0002】

【従来の技術】現在、電子メールの普及により、コンピュータ上で日本語入力処理を利用し文書を作成し、ネットワークを介し種々の情報のやり取りを行うケースが増えている。そのやり取りにおいて、文書内で使われている人名、地域名、専門用語あるいは文書内で固有に表現された単語／複合語等を繰り返し入力する場合が多い。

このような日本語入力処理では、単語辞書に含まれない単語／複合語があれば、文節の区切り直し機能等により単語／複合語を作り出すか、この操作で作り出すことができない場合は、単語登録する若しくは、アプリケーションの持つカットアンドペースト機能を用いて該単語／複合語を目的の位置に張り付けたりしている。

【0003】そして、このように単語辞書に登録されていない単語を自動的に学習あるいは登録する方法が、例えば特開平7-28808号公報及び特開平2-163874号公報に記載されている。これらは、いずれも利用者が一度必ず漢字変換操作を行い、その結果作り出した文字列の中で、単語辞書に未登録な単語があれば自動的に学習あるいは登録する方法であり、二度目の入力から有効である。

【0004】

【発明が解決しようとする課題】しかしながら、かかる従来の日本語入力処理方法にあつては、かな漢字変換する場合に、単語辞書に存在する文字で構成される単語／熟語しか、扱うことができないため、電子メールでやり取りを行う文書内に、他の単語辞書でユーザ登録された単語で、利用者が使用しているかな漢字変換では作成できない漢字が含まれている場合、それらの単語を再入力するには、単語登録機能を用いて利用者の単語辞書へ追加登録するか、アプリケーションの持つカットアンドペースト機能により張り付けるかしか方法がなく、その作業が面倒で、そのたびに作業が中断し、作業能率が低下するという課題があった。

【0005】また、電子メールでやり取りを行う文書内に、利用者の単語辞書に存在しない未登録語が存在し、それらの単語を再入力しようとする場合、文字列の変換を行った後、再度区切り直し等を行って単語を作り出す必要があるという課題があった。その理由は、単語を学習若しくは単語辞書へ登録するためには、その単語の読み及び品詞情報が必要だが、文書内の文字列からその情報を取得できないためである。

【0006】前記目的達成のため、この発明にかかる単語辞書に存在しない漢字等から構成される単語でも、最初の変換候補として表示することが可能となり、単語入力のために余分な操作を行わずに、思考を中断することなく日本語入力を行うことでできる日本語入力システムを得ることを目的とする。

【0007】

【課題を解決するための手段】前記目的達成のため、請求項1の発明にかかる日本語入力システムは、かな漢字変換部に設けられて、利用者が文書内に日本語入力を開始し、確定した文字列の中で、ユーザ登録単語及び単語辞書に存在しない未登録語を抽出する学習単語候補切出し部と、利用者の入力完了して文書を保存する際、実際に入力された文字列以外に前記学習単語候補切出し部で抽出したユーザ登録単語及び未登録語からなる学習単語

語情報も併せて文書ファイル内に追加して保存するファイル保存部と、前記文書ファイルが開かれた場合、そのファイル情報をかな漢字変換部へ通知するファイルオープン部と、前記かな漢字変換部に設けられて、前記ファイル情報の通知を受けて、文書ファイル内の前記学習単語情報をメモリ上へロードし、その学習単語情報に含まれる単語が現在使用中の単語辞書に存在するか否かを検索し、存在しない未登録の単語のみを取込む学習単語候補取込み部とを設けて、日本語変換処理部に、利用者が入力した読み情報が、前記学習単語候補取込み部で取込んだ単語に該当する場合、文法的に問題がなければ最初の変換候補として出力させ、更に、その単語が選択された場合は、単語登録させるかを利用者に確認させ、または必要なら自動的に単語登録を行わせるようにしたものである。

【0008】また、請求項2の発明にかかる日本語入力システムは、前記学習単語候補取込み部に、前記学習単語候補切出し部で切出した単語情報と現在使用中の単語辞書とを照合した結果、未登録語と断定した単語を第1の学習単語候補テーブルに格納させるようにしたものである。

【0009】また、請求項3の発明にかかる日本語入力システムは、前記かな漢字変換部に、前記学習単語候補取込み部で取込んだ単語を選択して単語辞書へ登録させる単語登録部を設けたものである。

【0010】また、請求項4の発明にかかる日本語入力システムは、前記かな漢字変換部に、前記日本語変換処理部で作成した変換結果を表示装置に表示させる画面表示部を設けたものである。

【0011】

【発明の実施の形態】以下、この発明の実施の一形態を図について説明する。図1において、110は利用者がキーボード装置113より入力したキー情報をもとに、変換処理に必要な読み情報を作成したり、また文字列の変換、文字列の選択等のかな漢字変換処理を判断する入力処理部、111はかな漢字変換処理の結果をCRTディスプレイなどの表示装置114へ出力する画面表示部、108は一般的な漢字、ひらがな、熟語や利用者が追加したユーザ単語が登録されている単語辞書、105は学習単語候補切出し部であり、利用者がキーボード装置113より読み文字列を入力し、変換を行った後、どのような操作が行われるか監視し、文節を区切り直すか、またはカタカナ変換（文節内の文字列をすべてカタカナ文字列に変更する）が行われた場合、その文節の位置を記憶しておき、利用者が変換文字列を確定した段階で、文節の位置をもとに未登録語を抽出し、また、確定文字列中にユーザ登録単語が含まれていた場合は、その文字列も抽出する。

【0012】更に、この学習単語候補切出し部105は付属語テーブルを用いて品詞情報及び自立語を決定し、

自立語部分とその読み文字列及びフラグ（品詞と学習単語候補であることを示す）情報を、第2の学習単語候補テーブル109に保存する。104は利用者が入力を完了し文書を保存する際、前記第2の学習単語候補テーブル109の情報を参照し、文書ファイル内に学習単語候補テーブル109として追加保存することが可能なファイル保存部である。101は利用者により前記情報が付加された文書ファイルがオープンされた場合、文書ファイル情報を学習単語候補取込み102に通知するファイルオープン部である。

【0013】また、102は前記ファイルオープン部101から入手したファイル情報をもとに、文書ファイル内の第2の学習単語候補テーブル部109のみをメモリ上へロードし、第2の学習単語候補テーブル109内の単語が既に単語辞書108に登録されていないか検索し、存在しない単語のみを第1の学習単語候補テーブル103へ保存する学習単語取込み部、106は入力処理部110から入手した情報をもとに単語辞書108を検索し、変換処理等を行う日本語変換処理部で、更に、利用者が入力した読みに該当する単語が第1の学習単語候補テーブル103に存在する場合はその単語を優先的に出力し、利用者がその文字列を確定した場合に単語辞書108へ登録するか確認後、必要なら自動的に単語登録を行う機能も有する。107は前記日本語変換処理部106より単語、単語の読み、品詞情報を入手し単語辞書108へ登録を行う単語登録部である。

【0014】次にこの発明の動作について、図2～図8を参照して詳細に説明する。図2はこの発明における特徴的な利用形態を示した図である。電子メールシステムにおいて各利用者A、Bは、個々に単語辞書108A、108Bを有しており、同一の日本語入力システムでも、運用の仕方（ユーザ登録単語や学習文字列の違い）により、単語辞書の内容が異なってくる。例えば、利用者Bは“古手川”という人名をユーザ単語として単語辞書108Bに登録しており、その文字列を含んだ文書を作成し利用者Aに送付すると、利用者Aは受け取った文書上で、再度古手側”を入力したいが、単語辞書108Aにはその単語が含まれていないため、区切り直し操作文字列を作る必要がある。あるいは、辞書に存在しない漢字が含まれている単語（利用者Bがユーザ登録した単語）は、かな漢字変換を用いた操作では入力が行えない。

【0015】この発明では、文書内に既に入力されている単語／熟語（漢字やカタカナで構成される）のうち、利用者の辞書に存在しない未登録語を他の利用者が文書作成時に抽出し、文書ファイル内に保持することで、その文書ファイルを再編集する場合にそれらの単語を抽出して自動的に学習することで、未登録語を意識することなく日本語入力が可能になる。

【0016】図3は前記文書ファイルの構造図で、この

文書ファイルは文書ヘッダ部301、画面表示文字列302および学習単語候補テーブル部303から構成されている。更に学習単語候補テーブル部303は学習単語候補テーブルの容量を示す全データ長304、格納されている単語／熟語の数を示す学習文字列数305、及び文字列登録エリア306を構成する読み文字列307、単語／熟語文字列としての学習文字列308、品詞と学習単語候補テーブル部303内に含まれる単語／熟語であることを示すフラグ309などから構成されている。図1に示す第1、第2の学習単語候補テーブル103、109も同様の構成を持つ。また、図4は品詞の決定及び自立語部分の判定に使用する付属語テーブルを示す。

【0017】図5は図1における学習単語候補切出し部105の詳細を示す機能構成図で、これが操作監視部401、学習単語候補検出部402、学習単語候補情報作成部403からなる。これらのうち、操作監視部401は利用者が読み文字列を入力し、日本語変換処理を行った後、その変換文字列に対し文節の区切り直しあるいは、カタカナ変換が行われるかを監視し、修正が行われた場合は、修正された文節を記憶する。

【0018】また、学習単語候補検出部402は利用者が修正を完了し確定した文字列に対し、操作監視部401で記憶した文節位置から、付属語が含まれる文字列が存在する文節までを、学習単語候補文字列として抽出する。更に、確定文字列内にユーザ登録単語が存在するかを、かな漢字変換処理時の内部情報を参照することで判断し、あれば抽出する。さらに学習単語候補情報作成部403は前記学習単語候補検出部402で抽出した文字列に対し、図4に示すような付属語テーブルを検証することで、品詞を決定し、更に文字列から付属語部分を取り除き、図1に示す第2の学習単語候補テーブル109へ格納する。

【0019】図6は図1における学習単語候補取込み部102の詳細を示す機能構成図で、これが文書情報ロード部501と学習単語候補テーブル検索部502とから構成されている。文書情報ロード部501は利用者が文書ファイルをオープンした時点で、図1に示すファイルオープン部101よりファイル情報を受け取り、文書ファイル内の学習単語候補テーブル部303をメモリ上へロードし、第1の学習単語候補テーブル103を作成する。学習単語候補テーブル検索部502は第1の学習単語候補テーブル103に含まれる単語について、単語辞書に存在するか否かを検証し、既に存在する単語があれば、第1の学習単語候補テーブル103から削除する。

【0020】また、図7は学習単語候補切出し部105における処理手順を示すフロー図である。まず、利用者は入力を開始する(ステップS601)。例えば、“大動物展が開催される”と入力すると、かな漢字変換処理が実行され(ステップS602)、“大同／物展が／開催される”と変換された場合の未登録語の抽出方法につ

いて説明する。いま、前記変換が正しく、利用者は文字列を確定しない場合には(ステップS603)、修正処理を監視し(ステップS604)、文字を区切り直すか、カタカナ変換を実行し(ステップS605)、変換文字列を確定したとき文節位置を記憶する(ステップS606)。すなわち、利用者は意図した変換文字列が得られないため、第1文節の“大同”を“大／同”で区切り直しを行う。その結果“大／動物／展が／開催される”との変換結果を得る。そして、このように区切り直しが行われた第1文節を記憶する。なお、ステップS605で文節変更やカタカナ変換が実行されない場合には、その他の処理を実行する(ステップS613)。

【0021】次に文字列を確定した場合には、利用者が意図した変換結果が得られたため、ステップS603で文字列を確定する。そして、記憶された文節があるか、更にユーザ登録単語が含まれるか否かを確認する(ステップS607)。なお、ユーザ登録単語が存在するかどうかは、変換処理時の内部情報を参照することで、容易に確認できる。続いて、修正のあった文字列(未登録語)およびユーザ登録単語を抽出し(ステップS608)、未登録語については、ステップS607で記憶した文節位置にもとづいて、その文節内の文字列から付属語テーブルに含まれる付属語が存在する文節までを学習単語候補文字列として抽出する。例えば、“大／動物／展が／開催される”の場合、第1文節から文字列をサーチし、付属語“が”が含まれる第3文節までの文字列“大動物展が”を抽出する。ユーザ登録単語については、全て1文節で構成されるため、該当文節の文字列を抽出する。

【0022】次にステップS608で抽出した文字列の付属語から、品詞を決定し(ステップS609)、更に、その付属語部分を取り除き単語を抽出する。“大動物展が”の場合、付属語テーブルより付属語“が”は名詞と判断でき、“が”を取り除いた“大動物展”を単語として抽出できる。続いて、抽出した“大動物展”が、既に学習単語候補として、第1および第2の学習単語候補テーブル103、109に存在するか否かを確認し、存在すれば該単語を学習単語候補から削除する(ステップS611)。もし、存在しなければ、取り出された単語を第2の学習単語候補テーブル109に格納する(ステップS612)。

【0023】図8は図1における学習単語候補取込み部102による処理手順を示すフロー図である。まず、図1におけるアプリケーション部114のファイルオープン部101よりファイル情報を入手し(ステップS701)、文書ファイル内の学習単語候補テーブル部303をメモリ上へロードし、第1の学習単語候補テーブル103を作成する(ステップS702)。さらに、単語辞書を検索し(ステップS703)、既に登録されている単語がないか確認する(ステップS704)。既に存在

する単語が含まれている場合は、該単語を削除し、第1の学習単語候補テーブル103を再編成する(ステップS705)。

【0024】図9は図1における日本語変換処理部106による処理手順を示すフロー図である。これによれば、まず、処理の種別を判定し(ステップS801)、変換処理である場合には、利用者が読み文字列を入力し変換処理を実行した際(ステップS802)、入力した読み文字列の中に第1の学習単語候補テーブル103に存在する読みが指定されているか否かを判断し(ステップS803)、存在した場合は品詞情報のチェックを行い、その読みに該当する単語が変換候補として成立するか否かを確認する(ステップS804)。成立すると判断した場合は、最初の変換候補として出力するよう変更する(ステップS805)。

【0025】更に、学習処理を選択し、利用者が変換文字列を確定した場合、その単語を選択したか否かを判断し(ステップS806)、選択された場合は単語辞書へ登録するか否かを利用者に確認する(ステップS807)。確認した結果、辞書への登録を要求された場合は、該単語及び単語の読み及び品詞情報を図1における単語登録部107へ通知し、単語辞書への登録を行う(ステップS808)。なお、ステップS806で第1の学習候補テーブルに含まれる候補がないと判定された場合には、従来の学習処理を実行する(ステップS809)。また、ステップS801でその他の処理が選択された場合には、既存の処理を実行する(ステップS810)。

【0026】

【発明の効果】以上のように、この発明によれば、既に入力済みの文字列以外に、その文字列内に含まれるユーザ登録単語及び未登録語の情報を文書ファイル内に持ち、その情報をもとに、事前に学習することができるようにしたので、既に日本語文字列が入力されている文書を利用者が編集／修正する際、その文書内の単語を再入力するケースで、現在使用している単語辞書に存在しない単語が含まれていても、かな漢字変換の最初の変換候補として出力することが可能になる。これにより利用者は余分な操作を行わず、思考を中断することなく文字の

入力を行うことができるという効果が得られる。

【0027】また、この発明によれば、文書ファイルの情報から自動学習した単語を利用者が入力した際、利用者の指示により、自動的に単語辞書へ登録することが可能になる。これにより、登録に必要な読みや品詞情報を予め備えることによって、利用者は単語登録機能を使って単語や読み文字列及び品詞情報の設定等煩わしい操作を行わなくとも、単語辞書へ登録することができるという効果が得られる。

【図面の簡単な説明】

【図1】 この発明による日本語入力システムの実施の一形態を示すブロック図である。

【図2】 図1における日本語入力システムの利用の形態を示す説明図である。

【図3】 この発明における文書ファイルの構成を示す構成図である。

【図4】 この発明における付属語テーブルの構成を示す説明図である。

【図5】 図1における学習単語候補切出し部の概略構成を示すブロック図である。

【図6】 図1における学習単語候補取込み部の概略構成を示すブロック図である。

【図7】 図1における学習単語候補切出し部の動作を示すフローチャートである。

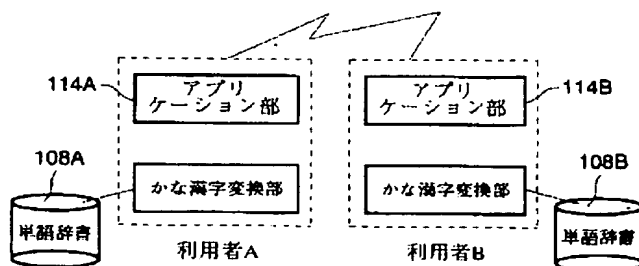
【図8】 図1における学習単語候補取出し部の動作を示すフローチャートである。

【図9】 図1における日本語変換処理部の動作を示すフローチャートである。

【符号の説明】

- 101 ファイルオープン部
- 102 学習単語候補取込み部
- 103 第1の学習単語候補テーブル
- 104 ファイル保存部
- 105 学習単語候補切出し部
- 106 日本語変換処理部
- 107 単語登録部
- 108 単語辞書
- 111 画面表示部
- 118 かな漢字変換部

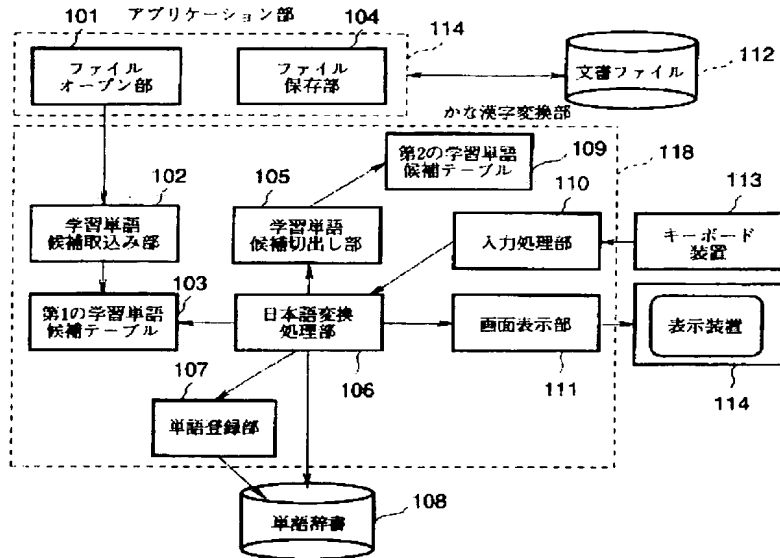
【図2】



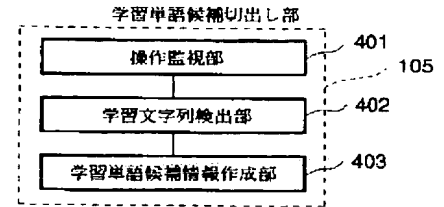
【図4】

品詞	付属語
名詞	が、か、の、に、を、は、も、と...
サ変名詞	をし、し、をさ、さ、をせ、せ、をする、する をすれ、すれ、をしろ、しろ、をせよ、せよ

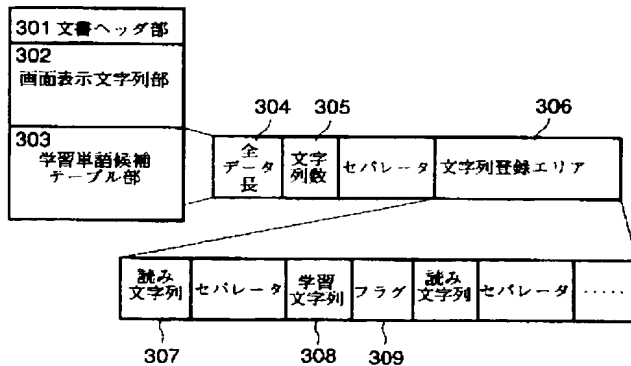
【図1】



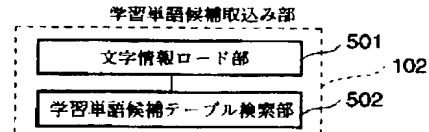
【図5】



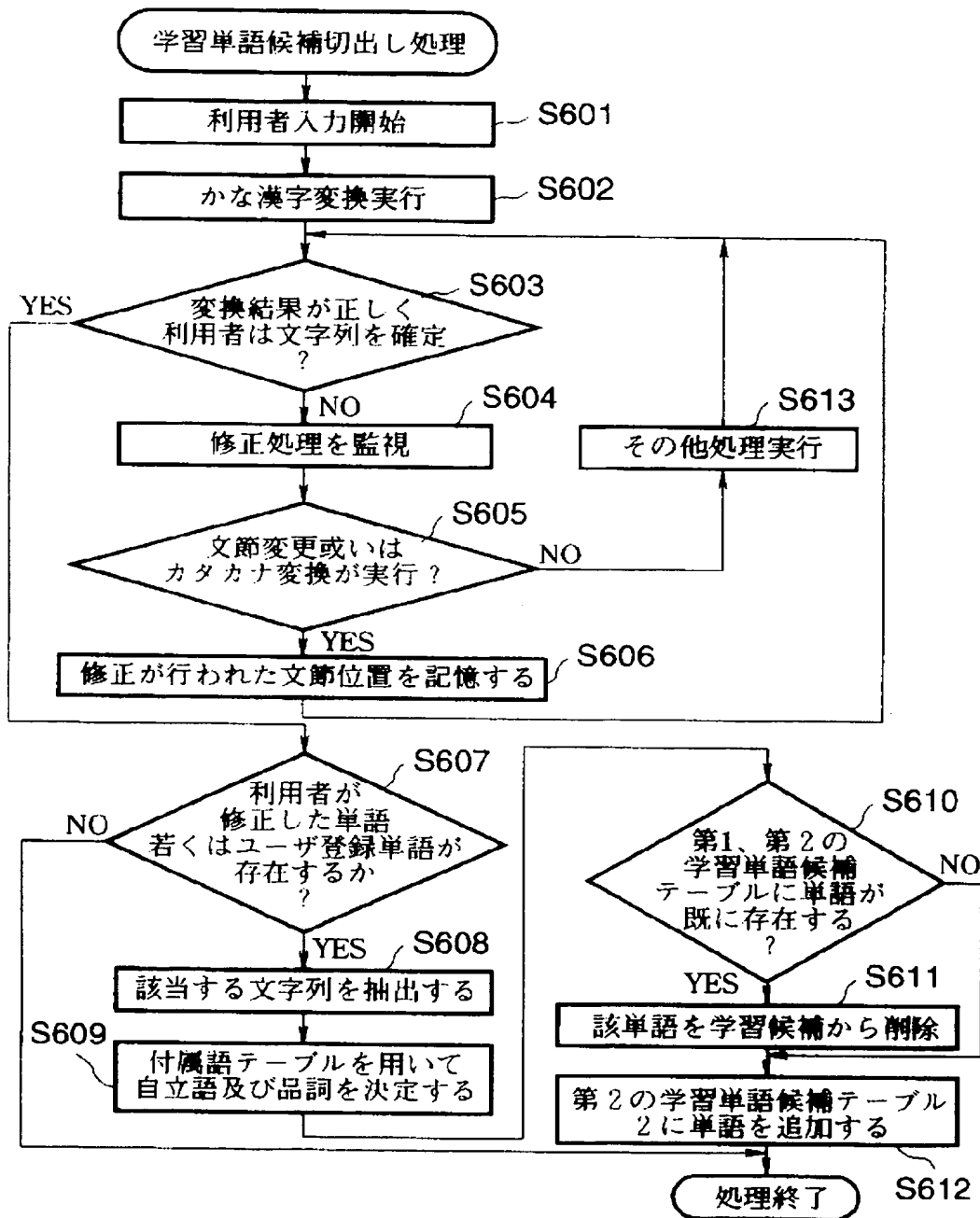
【図3】



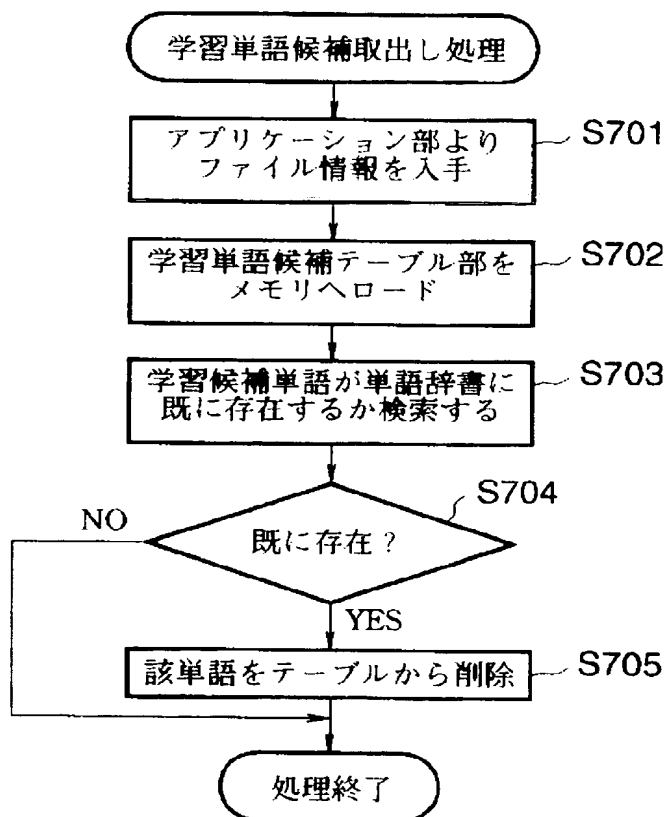
【図6】



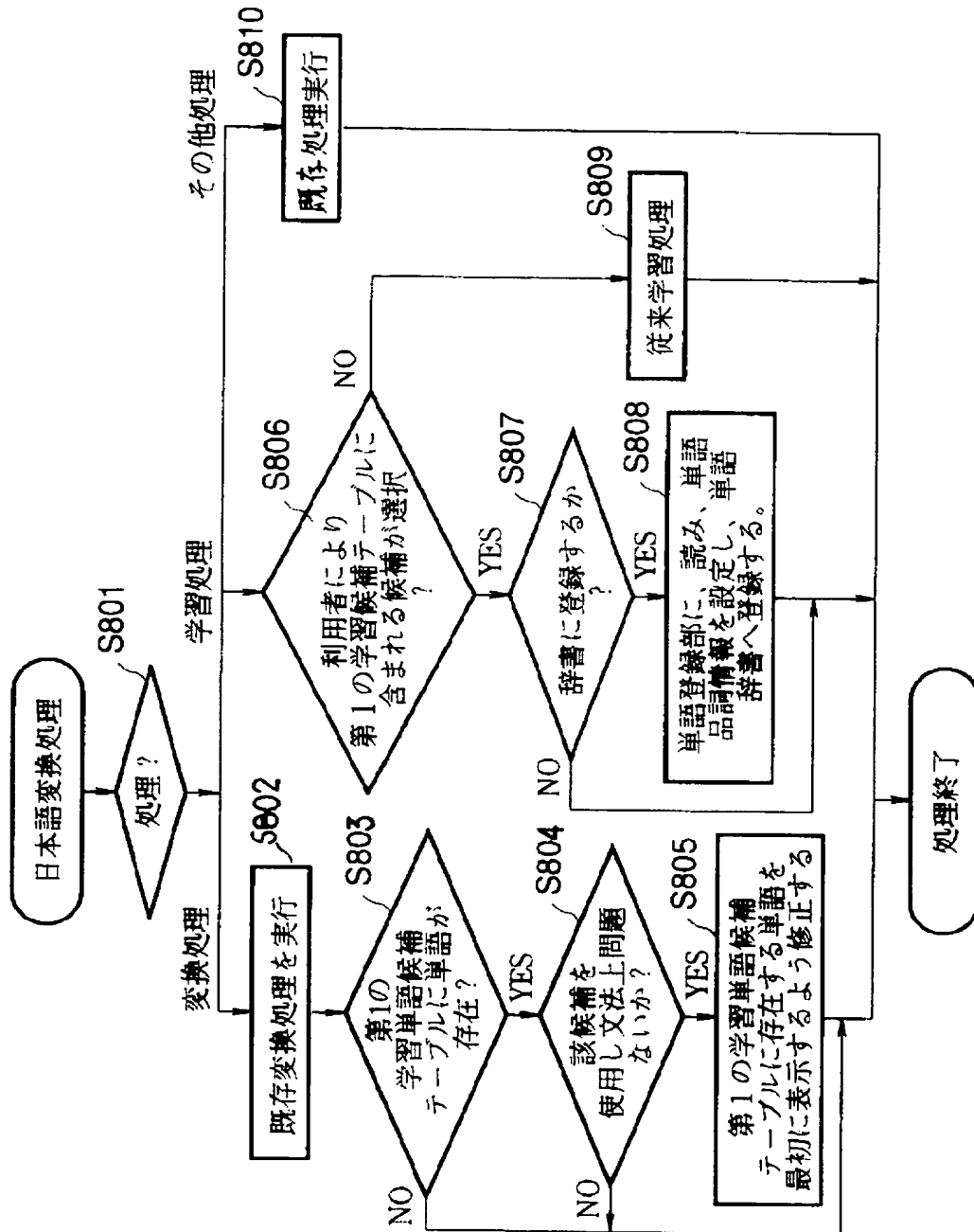
【図7】



【図8】



【図9】



THIS PAGE BLANK (USPTO)